

Implementasi Algoritma *Support Vector Machine* Dalam Klasifikasi Komentar Pengguna Produk Skintific di *E-Commerce*

DOI: <http://dx.doi.org/10.35889/jutisi.v14i2.3137>

Creative Commons License 4.0 (CC BY – NC)

Priti Rindi Artika^{1*}, Aidil Halim Lubis²

Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

*e-mail Corresponding Author: pritrindyartika@gmail.com

Abstract

The large number of consumer reviews of skincare products such as Skintific 5X Ceramide Barrier Repair Moisture Gel on e-commerce platforms raises the need for automated sentiment analysis. This study classifies 2,000 user comments from the Sociolla app using the Support Vector Machine (SVM) algorithm. Data were obtained through web scraping and processed through preprocessing, lexicon-based labeling, and word weighting using TF-IDF. SVM with a linear kernel was used to distinguish positive and negative comments. Performance evaluation using a confusion matrix resulted in an accuracy of 89.73%, a precision of 0.94, a recall of 0.75, and an F1-score of 0.83 for the positive class, and a precision of 0.88, a recall of 0.98, and an F1-score of 0.93 for the negative class. These results indicate that SVM is effective for sentiment classification in online beauty product reviews.

Keywords: E-commerce; Support Vector Machine; Sentiment Analysis; TF-IDF; Sociolla; Skintific; Lexicon-based

Abstrak

Banyaknya ulasan konsumen terhadap produk perawatan kulit seperti Skintific 5X Ceramide Barrier Repair Moisture Gel di platform e-commerce menimbulkan kebutuhan akan analisis sentimen otomatis. Penelitian ini mengklasifikasikan 2000 komentar pengguna dari aplikasi Sociolla menggunakan algoritma Support Vector Machine (SVM). Data diperoleh melalui web scraping dan diproses dengan tahapan preprocessing, pelabelan berbasis lexicon, serta pembobotan kata menggunakan TF-IDF. SVM dengan linear kernel digunakan untuk membedakan komentar positif dan negatif. Evaluasi performa menggunakan confusion matrix menghasilkan akurasi sebesar 89,73%, precision 0,94, recall 0,75, dan F1-score 0,83 untuk kelas positif, serta precision 0,88, recall 0,98, dan F1-score 0,93 untuk kelas negatif. Hasil ini menunjukkan bahwa SVM efektif untuk klasifikasi sentimen pada ulasan produk kecantikan secara daring.

Kata kunci: E-commerce; Support Vector Machine; Analisis Sentimen; TF-IDF; Sociolla; Skintific; Lexicon-based

1. Pendahuluan

Kehadiran ulasan pengguna pada platform e-commerce telah menjadi sumber informasi penting dalam proses pengambilan keputusan konsumen [1]. Ulasan tersebut merefleksikan pengalaman nyata yang dapat memberikan gambaran kualitas produk secara langsung [2]. Dalam konteks produk perawatan kulit, seperti Skintific 5X Ceramide Barrier Repair Moisture Gel, analisis sentimen terhadap komentar pengguna menjadi penting untuk memahami persepsi dan kepuasan konsumen terhadap produk, serta untuk mendukung strategi pemasaran dan pengembangan produk lebih lanjut [3].

Saat ini, produk Skintific dikenal luas di pasar Indonesia dan menduduki peringkat pertama dalam penjualan produk kecantikan di platform Sociolla. Namun, banyaknya komentar pengguna yang bersifat subjektif dan tidak terstruktur menyulitkan perusahaan maupun

konsumen lain dalam menyimpulkan secara objektif sentimen mayoritas. Masalah yang muncul adalah tidak tersedianya sistem otomatis yang dapat mengelompokkan komentar berdasarkan sentimen positif atau negatif secara akurat dan cepat. Hal ini berdampak pada keterlambatan perusahaan dalam menanggapi masukan konsumen serta menghambat analisis pasar yang lebih mendalam.

Solusi yang ditawarkan adalah penerapan algoritma *Support Vector Machine* (SVM) untuk klasifikasi sentimen pada komentar pengguna. SVM dikenal efektif dalam menangani klasifikasi teks, khususnya data tidak terstruktur seperti komentar atau ulasan. Dengan menggabungkan pendekatan *lexicon-based* untuk pelabelan data dan metode *TF-IDF* untuk pembobotan kata, algoritma SVM dapat membedakan sentimen secara optimal. Penelitian sebelumnya menunjukkan bahwa SVM memiliki akurasi tinggi dalam klasifikasi komentar konsumen, seperti pada studi oleh Larasati et al. (2024) yang mencapai akurasi hingga 98% dalam klasifikasi ulasan produk kecantikan di media sosial [4].

Tujuan dari penelitian ini adalah untuk membangun model klasifikasi sentimen berbasis SVM terhadap komentar pengguna produk *Skintific* di *platform Sociolla*. Manfaat dari penelitian ini adalah memberikan solusi analitik yang akurat dan efisien bagi pelaku industri kosmetik dalam memahami tanggapan konsumen secara sistematis, serta memberikan wawasan tambahan bagi pengembangan sistem analisis sentimen berbasis kecerdasan buatan di sektor *e-commerce*.

2. Tinjauan Pustaka

Penelitian oleh Ria Cantika Larasati et al. (2024) membahas analisis sentimen terhadap produk moisturizer di *platform Twitter* (X) menggunakan algoritma *Support Vector Machine* (SVM). Penelitian ini menggunakan pendekatan *supervised learning* dengan metode pelabelan manual, serta mengukur performa menggunakan parameter *accuracy*, *precision*, *recall*, dan *f1-score*. Hasilnya menunjukkan bahwa SVM mampu mengklasifikasikan sentimen dengan akurasi mencapai 98%, recall 100%, dan f1-score sebesar 99% terhadap komentar netral, positif, dan negatif [5].

B. A. Maulana et al. (2024) melakukan klasifikasi sentimen terhadap aplikasi investasi Pluang menggunakan algoritma *Naïve Bayes* dan SVM. Penelitian ini menggunakan data komentar pengguna dari Google Play Store, kemudian dilakukan *preprocessing* dan pembobotan menggunakan TF-IDF. SVM menunjukkan performa yang lebih baik dibandingkan *Naïve Bayes*, dengan akurasi mencapai 91,25%. Penelitian ini juga menyoroti pentingnya teknik balancing data karena terdapat ketimpangan jumlah data antar kelas [6].

F. Maulana Herza dan B. Rahmat (2024) membandingkan kinerja algoritma SVM dan *Logistic Regression* dalam menganalisis sentimen terhadap ulasan hotel Shangri-La Surabaya. Penelitian ini mengimplementasikan fitur seleksi menggunakan RFE (*Recursive Feature Elimination*) untuk mengurangi dimensi data. SVM memberikan hasil lebih akurat dalam membedakan komentar positif dan negatif dengan bantuan pembobotan TF-IDF [7].

Penelitian ini memiliki *state of the art* pada beberapa aspek dibandingkan penelitian sebelumnya. Pertama, berbeda dari studi sebelumnya yang sebagian besar menggunakan media sosial sebagai sumber data, penelitian ini memanfaatkan data ulasan dari *platform e-commerce* khusus produk kecantikan, yakni *Sociolla*, yang memiliki tingkat kontrol kualitas ulasan yang lebih baik. Kedua, proses pelabelan sentimen dilakukan dengan pendekatan *lexicon-based*, bukan manual, sehingga meningkatkan efisiensi dan konsistensi klasifikasi awal [8]. Ketiga, selain menggunakan kernel linear, penelitian ini juga menghitung *kernel polynomial* dan menguraikan perhitungan matriks Hessian serta pembaruan parameter α pada tahap training, yang jarang ditampilkan secara rinci dalam penelitian sejenis. Terakhir, penelitian ini menampilkan proses klasifikasi data uji hingga tahap akhir prediksi dan analisis kata (*wordcloud*), memberikan gambaran menyeluruh dari *preprocessing* hingga interpretasi hasil. Dengan demikian, *novelty* dari penelitian ini terletak pada kombinasi pendekatan teknis yang lebih komprehensif, objek yang lebih spesifik, dan metode pelabelan otomatis berbasis leksikon untuk mendukung klasifikasi komentar secara efisien dan terukur.

3. Metodologi

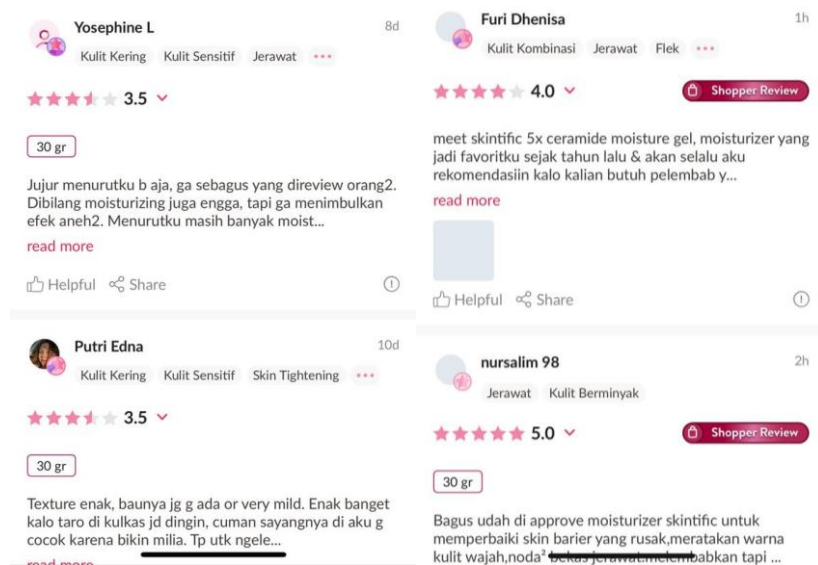
3.1. Data dan Teknik Pengumpulan Data

Peneliti menggunakan 2000 data komentar yang dimana berisi nama pengguna, rating dan ulasan. Dalam pengumpulan data dari *website sociolla* dibantu dengan *Selenium* dan

BeautifulSoup untuk melakukan *scraping*. Proses diawali dengan inisialisasi *WebDriver* untuk membuka halaman target, yang kemudian dilanjutkan dengan menutup *pop-up* jika muncul. Konten halaman dimuat secara menyeluruh dengan manfaat fungsi *scroll* otomatis. Data yang dikumpulkan diekstraksi dari elemen HTML menggunakan *BeautifulSoup* dan disimpan dalam file CSV.

Tabel 1. Data Ulasan Pengguna Produk *Skintific 5x Ceramide Barrier Repair Moisture Gel*

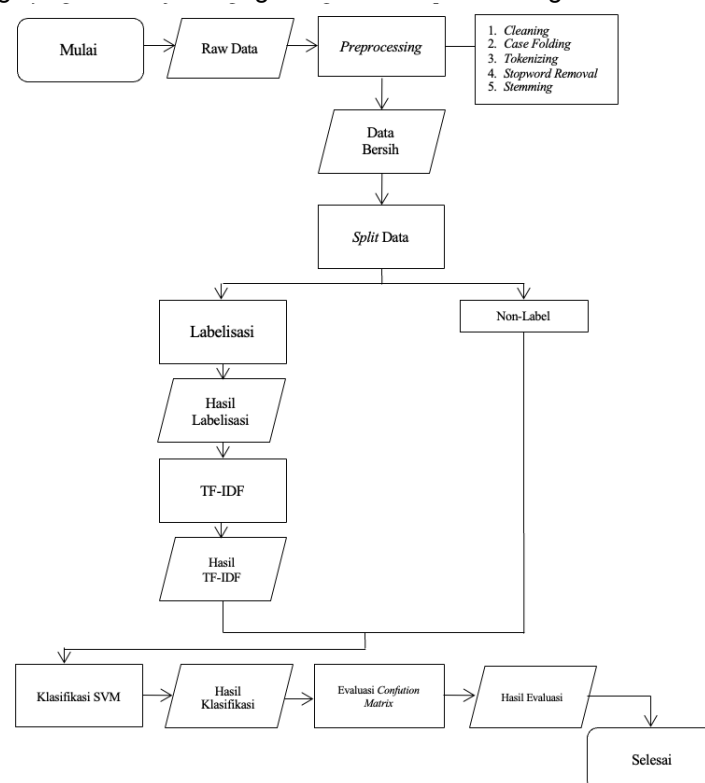
No	Pengguna	Rating	Ulasan
1	Fatima Benyaich	5	please suka bgt??? beneran enak banget gitu loh pas dipake 🥰❤️ kulitku tuh bener2 yang kering bgt sampe bingung mau pake apa lagi. tapi pas nyobain produk ini tuh langsung cocok gitu loohhh huhuuu cinta bgt 🥰❤️
2	Hiel Hiel	1	b.o parah kulit aku make ini mungkin kalo tipe kulit yang dry skin cocok, waktu peetama nyoba bikin kusam parah habis itu gayal atau cekat cekit di beberapa bagian dan merah merah aku langsung stop dan akhirnya aku make nya buat pelembab lutut smaa siku doang mana mahal lagi :)
3	Ajrina Ziezie	5	sesukaaa itu, udah abis 2 jar dan sudah repurchase lagi untuk jar yang ketiga, dari segi teksturnya gel berwarna putih yang cepat menyerap, tetapi tetep bikin lembab sampai pagi, engga greasy sama sekali, bikin masalah dikulit jadi cepet calm, wanginya tidak mengganggu malah cenderung bikin tenang, suka banget! yah minusnya di packaging aja sih, kurang travel friendly, dan aku personally kurang suka sama skincare dengan packaging jar berspatula gitu, tapi selain packaging semuanya very recommended !!!
4	Aisa Putri	1	baru 2x pake muka langsung BREAKOUT!! semua jerawat langsung muncul trus bukannya keluar malah jd mendem" gtuu, bikin susah ilang, kirain purging, gua coba pake lagi, malah muncul lagi, dan makin banyakk. padahal udh beli rangkaian skintific dan gacocok semua.
5	Rizkah Anggarani	5	jar keduaaaa, penyelamat saat kulit flare up, breakout, damage skin barrier, auto langsung balik ke moist ini, pake essence toner yang 4D hyaluron, sama sunscreen and done. semingguan kulit udah kaleman dan progress banget. ow btw ini ada wangi tapi dari essential oil
...	...	...	...
10	Gita Aplrylia	4	muka aku merah - merah dan kulitnya kerasa tipis banget, tiap abis make up itu makin jerawat dan berkerak dan baru pakai akhir desember 2022. WOW! january ini jerawat ku udah ga ada lagi, dan kemerahan dimuka ku berkurang banyakkkk banget sampe aku kaget sendiri.



Gambar 1. Review Para Pengguna Produk *Skintific 5x Ceramide Barrier Repair Moisture Gel*

3.2. Prosedur Kerja Algoritma SVM

Langkah-langkah secara umum Algoritma SVM dalam mengklasifikasi komentar:



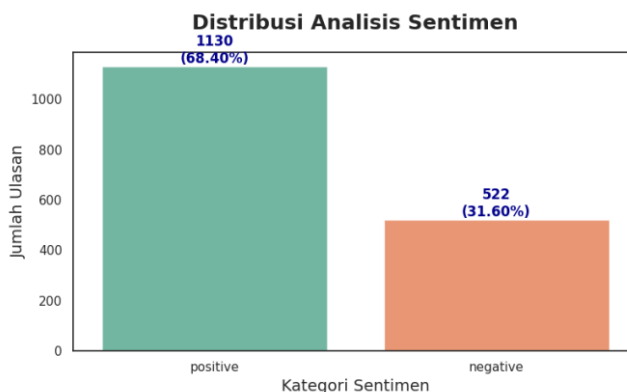
Gambar 2. Langkah-langkah Algoritma SVM

4. Hasil dan Pembahasan

4.1 Analisis Data

Data sentimen di kumpulkan melalui *library* python untuk melakukan web scraping dalam pengumpulan data ulasan produk dari *website Sociolla*. Data sentimen kemudian diubah ke dalam ekstensi *.csv* dan diolah menggunakan bahasa pemrograman python. Jumlah data sentimen yang dikumpulkan adalah 2000 data sentimen. Gambar 3. menampilkan hasil distribusi sentimen dari dataset ulasan pengguna produk Skintific 5X Ceramide Barrier Repair Moisture Gel

pada *platform Sociolla*. Dari total 1.652 ulasan yang telah melalui proses pelabelan, 1.130 ulasan (68,40%) tergolong sentimen positif, sedangkan 522 ulasan (31,60%) tergolong sentimen negatif. Pelabelan sentiment dilakukan dengan menggunakan metode *lexicon based* merupakan pendekatan analisis sentimen yang menentukan polaritas (positif, negatif, atau netral) [9], [10], [11], [12], [13], [14].



Gambar 3. Kategori Sentiment

4.2 Klasifikasi *Support Vector Machine*

4.2.1 Menghitung Kernel

Pada tahap klasifikasi, algoritma *Support Vector Machine* (SVM) diterapkan dengan menggunakan kernel linear, karena karakteristik data yang dianalisis menunjukkan bahwa data bersifat linier atau dapat dipisahkan dengan garis lurus. Kernel linear bekerja dengan menghitung dot product antar vektor dalam ruang fitur, yang kemudian dimanfaatkan untuk mengukur kedekatan atau hubungan antar data [15]. Tabel 2. menampilkan matriks kernel linear yang dihitung dari lima dokumen contoh (D1–D5) setelah teks ulasan diberi bobot menggunakan TF-IDF dan dinormalisasi. Kernel linear bekerja dengan menghitung *dot product* antar vektor dalam ruang fitur, yang kemudian dimanfaatkan untuk mengukur kedekatan atau hubungan antar data. Hasil dari perhitungan ini digunakan untuk membentuk hyperplane sebagai batas optimal pemisah antara kelas sentimen positif dan negatif. Perhitungan kernel linear ini mengikuti formulasi matematis tertentu yang menjadi dasar utama dalam mendukung kinerja proses klasifikasi secara optimal.

Tabel 2. Kernel Linear

	D1	D2	D3	D4	D5
D1	1.0000	0.0926	0.1504	0.0384	0.0245
D2	0.0926	1.0000	0.0210	0.0460	0.0664
D3	0.1504	0.0210	1.0000	0.1086	0.0159
D4	0.0384	0.0460	0.1086	1.0000	0.1563
D5	0.0245	0.0664	0.0159	0.1563	1.0000

Setelah memperoleh nilai kernel antar dokumen, tahap selanjutnya adalah menghitung matriks Hessian sebagai bagian dari proses pelatihan model *Support Vector Machine*. Sebelum perhitungan dilakukan, perlu ditetapkan beberapa parameter penting, seperti nilai α_i , C , γ , λ , serta jumlah maksimum iterasi. Masing-masing parameter ini memiliki peran strategis dalam membentuk matriks Hessian yang akan digunakan dalam proses klasifikasi sentimen, khususnya ketika menggunakan kernel polynomial yang membutuhkan pemetaan data ke ruang berdimensi lebih tinggi.

Tabel 3. Parameter dan Nilai

Parameter	Nilai
α_i	0
C	1
γ	0.1
λ	0.5
iterasi	5

4.2.2 Menghitung Matriks Hessian

Hasil komputasi matriks Hessian ditampilkan pada tahap ini sebagai bagian dari proses klasifikasi sentimen. Matriks tersebut memuat nilai-nilai yang merepresentasikan interaksi antar data dalam ruang fitur, yang diperoleh berdasarkan kernel polynomial yang digunakan. Informasi dalam matriks Hessian ini berperan penting dalam proses optimasi, khususnya dalam menentukan hyperplane pemisah yang paling optimal pada penerapan algoritma *Support Vector Machine* (SVM) dalam penelitian ini.

Tabel 4. Menghitung Matriks Hessian

	D1	D2	D3	D4	D5
D1	1.2500	0.3426	0.4004	-0.2884	-0.2745
D2	0.3426	1.2500	0.2710	-0.2960	-0.3164
D3	0.4004	0.2710	1.2500	-0.3586	-0.2659
D4	-0.2884	-0.2960	-0.3586	1.2500	0.4063
D5	-0.2745	-0.3164	-0.2659	0.4063	1.2500

4.2.3 Sequential Training SVM

Setelah diperoleh nilai-nilai dari matriks Hessian, tahapan selanjutnya adalah melakukan pelatihan model secara berurutan atau *sequential training*. Pada tahap ini, digunakan persamaan matematis tertentu sebagai dasar perhitungan untuk memperbarui nilai parameter α secara bertahap. Melalui pendekatan ini, model SVM belajar dari data latih dengan menyesuaikan bobot parameter secara iteratif hingga mencapai kondisi konvergen. Hasil dari proses ini diharapkan mampu membentuk hyperplane optimal yang berfungsi memisahkan kelas sentimen positif dan negatif dengan akurasi yang tinggi. Adapun rumus yang digunakan dalam proses pembaruan parameter disajikan sebagai berikut:

$$E_i = \sum_{j=1}^n \alpha_j D_{ij} \quad (1)$$

Keterangan:

E_i : Nilai error untuk data ke-i
 α_j : Nilai alpha untuk data ke-j
 H_{ij} : Elemen ke-ij dari matriks Hessian
 n : Jumlah data

Tahapan berikutnya dalam proses pelatihan adalah menghitung perubahan nilai α , yang dikenal sebagai $\delta\alpha$ (delta alpha). Perhitungan ini bertujuan untuk mengetahui sejauh mana nilai α perlu diperbarui pada setiap iterasi. Adapun rumus yang digunakan untuk menentukan nilai $\delta\alpha$ disajikan sebagai berikut:

$$\delta\alpha_i = \min\{\max[\gamma(1 - E_i), -\alpha_i], C - \alpha_i\} \quad (2)$$

Keterangan:

$\delta\alpha_i$: delta alfa ke-i
 γ : learning rate
 E_i : error rate
 α_i : alfa ke-i
 C : variable slack

Setelah memperoleh nilai $\delta\alpha$, langkah selanjutnya adalah menghitung nilai α yang diperbarui. Perhitungan ini dilakukan menggunakan persamaan tertentu yang telah ditetapkan untuk memperbarui parameter α berdasarkan hasil perubahan yang telah dihitung sebelumnya.

$$\alpha_i = \alpha_i + \delta\alpha$$

Tabel 5. α iterasi 1

No	Nilai
1	0,1
2	0,1
3	0.1
4	0.1
5	0,1

Setelah diperoleh nilai α pada iterasi pertama, proses dilanjutkan dengan menghitung nilai α untuk iterasi berikutnya menggunakan rumus yang sama, yaitu dengan memanfaatkan nilai $\delta\alpha$ yang telah dihitung sebelumnya. Proses ini dilakukan secara bertahap untuk memperbarui parameter model secara berkelanjutan. Nilai α akan terus disesuaikan pada setiap iterasi hingga mendekati kondisi optimal. Dalam penelitian ini, proses iterasi dilakukan hingga lima kali (5 iterasi) untuk memperoleh hasil klasifikasi yang stabil dan akurat.

Tabel 6. Menghitung nilai α

No	Nilai
1	0,3780
2	0,3925
3	0,3890
4	0,4293
5	0,4223

Berdasarkan nilai alpha (α) yang diperoleh setelah iterasi ke-5, Support Vector dapat diidentifikasi. Sebuah data dinyatakan sebagai Support Vector jika nilai α dari data tersebut lebih besar dari nol ($\alpha_i > 0$).

Tabel 7. Support Vector

Dokumen	Nilai Alpha (α_i) Akhir	Syarat ($\alpha_i > 0$)	Status
D1	0.3780	$0.3780 > 0$	Support Vector
D2	0.3925	$0.3925 > 0$	Support Vector
D3	0.3890	$0.3890 > 0$	Support Vector
D4	0.4293	$0.4293 > 0$	Support Vector
D5	0.4223	$0.4223 > 0$	Support Vector

Setelah proses iterasi selesai dilakukan, langkah selanjutnya adalah menghitung fungsi kernel untuk masing-masing kelas dengan merujuk pada nilai α tertinggi yang diperoleh. Berdasarkan hasil perhitungan hingga iterasi ke-5, diketahui bahwa nilai α kelas positif terbesar dimiliki oleh dokumen D2 dengan nilai $\alpha = 0,3925$ dan nilai α kelas positif diwakili oleh dokumen D4 dengan $\alpha = 0,4293$. Nilai tertinggi dari masing-masing kelas tersebut kemudian digunakan dalam perhitungan kernel, yang diambil dari elemen-elemen terkait pada matriks Hessien. Perhitungan untuk x^+ (Target: D2):

$$f(x^+) = \sum_{i=1}^5 \alpha_i y_i K(x_i, D_2) \quad (3)$$

$$D1: 0.3780 \times 0.0926 = 0.0350$$

$$D2: 0.3925 \times 1.0000 = 0.3925$$

$$D3: 0.3890 \times 0.0210 = 0.008169 \approx 0.0082$$

$$D4: -0.4293 \times 0.0460 = -0.0197478 \approx -0.0197$$

$$D5: -0.4223 \times 0.0664 = -0.0280277 \approx -0.0280$$

$$\text{Total: } f(x^+) = 0.0350 + 0.3925 + 0.0082 - 0.0197 - 0.0280 = 0.3880$$

Perhitungan untuk x^- (Target: D4)

$$f(x^-) = \sum_{i=1}^5 \alpha_i y_i K(x_i, D_4) \quad (4)$$

$$= (0.3780 \times 0.0384) + (0.3925 \times 0.0460) + (0.3890 \times 0.1086) + (-0.4293 \times 1.0000) + (-0.4223 \times 0.1563)$$

$$= 0.0145 + 0.0181 + 0.0422 - 0.4293 - 0.0660 = -0.4205$$

Tabel 8. Nilai x^+ dan x^-

x^+	x^-
0.3880	-0.4205

Setelah memperoleh hasil perhitungan nilai $K(x_i, x^+)$ dan $K(x_i, x^-)$, tahap selanjutnya adalah menentukan nilai bias (b). Perhitungan bias didasarkan pada persamaan yang telah ditetapkan sebelumnya, dengan mengacu pada hasil perhitungan kernel dan data support vector yang telah diidentifikasi pada tahap sebelumnya dalam penelitian ini.

$$b = -\frac{1}{2} [\sum_{i=1}^n \alpha_i y_i K(x_i, X^+) + \sum_{i=1}^n \alpha_i y_i K(x_i, X^-)] \quad (5)$$

Keterangan:

α_i : alfa ke-i

y_i : kelas data ke-1

$K(x_i, x^+)$: kernel kelas positif

$K(x_i, x^-)$: kernel kelas negatif

Dari persamaan diatas, maka diperoleh nilai b sebagai berikut:

$$b = -\frac{1}{2} [0.3880 + (-0.4205)]$$

$$b = -\frac{1}{2} [0.3880 - 0.4205]$$

$$b = -\frac{1}{2} [-0.0325]$$

$$b = 0.01625$$

Dengan demikian, nilai bias yang diperoleh adalah 0,01625, yang selanjutnya akan digunakan dalam fungsi klasifikasi untuk menentukan letak data uji terhadap hyperplane pemisah.

4.2.4 Pengujian SVM

Klasifikasi Data Uji 1:

Data Uji 1: ['berkalikali', 'repurchase', 'aman', 'acne', 'prone', 'skin', 'calming', 'mudah', 'serap', 'redness', 'reda', 'kulit', 'kering', 'kelupas', 'aman']

Langkah awal dalam memproses data uji adalah mengubahnya menjadi vektor numerik menggunakan skema pembobotan TF-IDF. Pada tahap ini, terdapat sebuah prinsip fundamental yang harus ditegakkan, yaitu konsistensi ruang fitur (*feature space*).

Model SVM yang telah dilatih hanya "memahami" fitur-fitur yang dipelajarinya selama proses pelatihan. Dalam konteks analisis teks, setiap kata unik dalam vokabulari data latih berfungsi sebagai satu fitur atau dimensi dalam ruang vektor. Oleh karena itu, vokabulari ini bertindak sebagai sebuah kamus atau cetak biru (blueprint) yang mendefinisikan struktur model.

Ketika data uji baru diperkenalkan, data tersebut harus diproyeksikan ke dalam ruang fitur yang sama persis. Artinya, kita hanya menghitung bobot untuk kata-kata dari data uji yang terdapat di dalam kamus data latih. Kata-kata yang tidak ada di dalam kamus tersebut (dikenal sebagai *Out of Vocabulary* atau OOV) secara metodologis harus diabaikan. Hal ini dikarenakan model tidak memiliki informasi statistik (seperti nilai IDF) mengenai kata-kata OOV tersebut dan tidak ada dimensi yang dialokasikan untuknya. Mengabaikan kata-kata OOV memastikan bahwa vektor data uji memiliki dimensionalitas yang konsisten dengan vektor data latih, sehingga operasi matematis dalam fungsi klasifikasi dapat berjalan secara valid.

Mudah: TF = 1, IDF = 0.6990 → Bobot = $1 \times 0.6990 = 0.6990$

Kulit: TF = 1, IDF = 0.3979 → Bobot = $1 \times 0.3979 = 0.3979$

Tabel 9. Nilai TF-IDF

Term	TF	DF	IDF	TF-IDF
mudah	1	1	0.6990	0.6990
kulit	1	1	0.3979	0.3979

Setelah mendapatkan representasi vektor TF-IDF, tahap selanjutnya adalah normalisasi vektor. Tujuan utama dari normalisasi adalah untuk menyeragamkan skala nilai dari setiap vektor dokumen. Dalam konteks TF-IDF, dokumen yang lebih panjang cenderung memiliki nilai bobot yang lebih besar, bukan karena isinya lebih relevan, tetapi hanya karena memiliki lebih banyak kata. Normalisasi menghilangkan bias yang disebabkan oleh panjang dokumen ini.

$$|v| = \sqrt{(0.6990)^2 + (0.3979)^2}$$

$$|v| = \sqrt{0.4886 + 0.1583}$$

$$|v| = \sqrt{0.6469}$$

$$|v| = 0.8043$$

Tabel 9. menampilkan hasil evaluasi *confusion matrix* terhadap model klasifikasi SVM yang digunakan dalam analisis sentimen terhadap data pengguna produk *skintific*. Berikut adalah nilai-nilainya:

Tabel 9. Confusion Matrix Model SVM

Predicted	Positive	Negative	Total
Positive	85 (TP)	29 (FN)	114
Negative	5 (FP)	212 (TN)	217
Total	90	241	331

Keterangan:

True Positive (TP) = 85, Data positif yang diklasifikasikan benar sebagai positif.

True Negative (TN) = 212, Data negatif yang diklasifikasikan benar sebagai negatif.

False Positive (FP) = 5, Data negatif yang salah diklasifikasikan sebagai positif.

False Negative (FN) = 29, Data positif yang salah diklasifikasikan sebagai negatif.

Untuk mengukur performa model klasifikasi sentimen yang digunakan dalam penelitian ini, dilakukan evaluasi menggunakan beberapa metrik umum dalam klasifikasi biner, yaitu akurasi, precision, recall, dan F1-score. Perhitungan dilakukan berdasarkan hasil confusion matrix yang telah diperoleh dari pengujian model SVM terhadap 332 data uji. Berikut adalah rumus dari masing-masing metrik:

1. Akurasi

$$Akurasi = \frac{TP + TN}{Total\ Data} = \frac{85 + 212}{331} = \frac{297}{331} = 89.72\%$$

2. Precision positif

$$Akurasi = \frac{TP}{TP + FP} = \frac{85}{85 + 5} = \frac{85}{90} = 94.44\%$$

3. Precision negative

$$Akurasi = \frac{TN}{TN + FN} = \frac{212}{212 + 29} = \frac{212}{241} = 87.96\%$$

4. Recall positif

$$Akurasi = \frac{TP}{TP + FN} = \frac{85}{85 + 29} = \frac{85}{114} = 74.56\%$$

5. Recall negative

$$Akurasi = \frac{TN}{TN + FP} = \frac{212}{212 + 5} = \frac{212}{217} = 97.65\%$$

6. F1 Score Positif

$$Akurasi = 2x \frac{0.9444 \times 0.7456}{0.9444 + 0.7456} = \frac{0.7041}{1.6900} = 83.33\%$$

7. F1 Score Negatif

$$Akurasi = 2x \frac{0.8797 \times 0.9765}{0.8797 + 0.9765} = \frac{1.7176}{1.8562} = 92.54\%$$

Tabel 10. Confusion Matrix Model SVM

Sentimen	Precision	Recall	F1 Score	Total
Positif	0.94	0.75	0.83	114
Negatif	0.88	0.98	0.93	217
	Akurasi		0.90	331
Macro Avg	0.91	0.86	0.88	331
Weighted Avg	0.90	0.90	0.89	331

Proses evaluasi ini menggunakan beberapa metrik yang umum dipakai dalam permasalahan klasifikasi biner, yaitu akurasi, *precision*, *recall*, dan *f1-score*. Evaluasi model berbasis *confusion matrix* memberikan gambaran yang jelas mengenai kinerja algoritma dalam mengklasifikasikan sentimen positif dan negatif. *Confusion matrix* menyajikan secara ringkas jumlah prediksi benar maupun salah yang dihasilkan oleh model untuk setiap kelas, sehingga memungkinkan analisis lebih mendalam terhadap kekuatan dan kelemahan model yang digunakan.

Hasil evaluasi terhadap model SVM dalam penelitian ini menunjukkan akurasi sebesar 89,73%, dengan nilai *precision*, *recall*, dan *F1-score* yang bervariasi untuk setiap kelas. Kinerja ini sejalan dengan temuan penelitian sebelumnya. Misalnya, Larasati et al. (2024) memperoleh akurasi hingga 98% dalam klasifikasi ulasan produk kecantikan di media sosial menggunakan SVM. Demikian pula, Maulana et al. (2024) dalam penelitian analisis sentimen terhadap aplikasi investasi *Pluang* menunjukkan bahwa SVM lebih unggul dibanding *Naïve Bayes*, dengan akurasi mencapai 91,25%. Hasil yang serupa juga ditunjukkan oleh Herza & Rahmat (2024), di mana SVM mampu memberikan hasil lebih baik dibanding *Logistic Regression* dalam klasifikasi ulasan hotel [7].

Dengan membandingkan hasil penelitian ini dengan studi-studi terdahulu, dapat disimpulkan bahwa algoritma SVM secara konsisten menunjukkan performa yang andal dalam analisis sentimen, baik pada data ulasan produk kecantikan, aplikasi digital, maupun ulasan layanan. Hal ini memperkuat validitas penggunaan SVM dalam konteks penelitian ini, sekaligus menegaskan bahwa capaian akurasi hampir 90% sudah berada dalam rentang kinerja optimal untuk klasifikasi teks.

4.5 Pembahasan

Hasil pengujian model klasifikasi sentimen dengan algoritma *Support Vector Machine* (SVM) pada penelitian ini menunjukkan akurasi sebesar 89,73%, dengan *precision* dan *recall* yang relatif tinggi, khususnya pada kelas negatif. Nilai ini memperlihatkan bahwa SVM mampu melakukan pemisahan kelas secara optimal meskipun data ulasan yang digunakan berasal dari domain spesifik, yaitu produk kecantikan di *platform e-commerce Sociolla*.

Jika dikaitkan dengan penelitian terdahulu, capaian penelitian ini sejalan dengan temuan Larasati et al. yang menunjukkan akurasi hingga 98% pada analisis sentimen produk kecantikan di media sosial menggunakan SVM. Selain itu, Maulana et al. juga menemukan bahwa SVM memberikan performa lebih baik dibandingkan *Naïve Bayes* dalam analisis sentimen terhadap aplikasi *Pluang*, dengan akurasi mencapai 91,25%. Hasil penelitian Herza & Rahmat pun memperkuat hal ini, di mana SVM unggul dibanding *Logistic Regression* dalam mengklasifikasikan ulasan hotel.

Kontribusi penelitian ini terhadap kumpulan temuan terdahulu dapat dilihat dari beberapa aspek. Pertama, penelitian ini menggunakan data ulasan dari *platform e-commerce* khusus produk kecantikan, yang berbeda dari sebagian besar penelitian sebelumnya yang menggunakan data media sosial atau aplikasi digital. Dengan demikian, penelitian ini memperluas cakupan penerapan SVM ke domain yang lebih spesifik, serta memberikan bukti bahwa SVM tetap robust meskipun diterapkan pada konteks data yang berbeda. Kedua, penelitian ini mengimplementasikan metode *lexicon-based labeling* dalam tahap pelabelan data. Pendekatan ini meningkatkan efisiensi dan konsistensi dalam proses anotasi dibandingkan metode manual yang banyak digunakan pada penelitian terdahulu. Hal ini menegaskan bahwa integrasi *lexicon-based labeling* dengan SVM dapat menjadi alternatif yang efektif dalam klasifikasi ulasan produk di *e-commerce*. Ketiga, kontribusi lainnya adalah pada aspek transparansi proses klasifikasi. Penelitian ini tidak hanya melaporkan hasil akhir berupa akurasi, tetapi juga menguraikan tahapan komputasi kernel, pembentukan matriks *Hessian*, hingga *sequential training*. Detail ini jarang disajikan dalam penelitian sebelumnya, sehingga penelitian ini memberikan pemahaman lebih komprehensif mengenai mekanisme internal SVM dalam memproses data teks.

Dengan demikian, penelitian ini tidak hanya menguatkan temuan sebelumnya bahwa SVM efektif untuk klasifikasi sentimen, tetapi juga memperluas aplikasi pada domain baru (ulasan produk kecantikan di *e-commerce*) dan menambahkan kontribusi metodologis melalui penggunaan *lexicon-based labeling* serta transparansi perhitungan matematis. Integrasi hasil penelitian ini dengan temuan-temuan terdahulu seperti pada [16 – 17] memperkaya literatur dalam bidang analisis sentimen berbasis *machine learning*, serta membuka peluang pengembangan lebih lanjut, misalnya melalui optimisasi parameter SVM atau integrasi dengan metode *word embedding* modern seperti Word2Vec dan BERT untuk meningkatkan pemahaman konteks linguistik yang kompleks.

5. Simpulan

Berdasarkan hasil penelitian mengenai Klasifikasi Komentar Pengguna Produk Skintific di *E-Commerce* dengan menggunakan algoritma *Support Vector Machine* (SVM), maka diperoleh beberapa kesimpulan sebagai berikut Penerapan Algoritma SVM dalam Analisis Sentimen,

Algoritma SVM berhasil diimplementasikan untuk melakukan klasifikasi komentar terhadap 2.000 review pengguna produk skintific di *e-commerce*. Proses implementasi mencakup tahapan pra-pemrosesan (*cleaning, tokenizing, stopword removal, stemming*), pembobotan fitur menggunakan TF-IDF, serta pelabelan data berbasis *lexicon*. Hasil klasifikasi menunjukkan bahwa SVM mampu membedakan opini pengguna menjadi dua kelas utama: positif dan negatif. Visualisasi seperti *wordcloud* dan grafik distribusi turut mendukung keberhasilan klasifikasi sentimen, yang mencerminkan opini nyata pengguna produk skintific. Akurasi Model SVM dalam Analisis Sentimen, berdasarkan evaluasi terhadap 332 data uji, model SVM menunjukkan performa yang cukup baik dengan akurasi sebesar 89.73%, *precision* 0,88 dan *recall* 0,98 untuk sentimen negatif, serta *precision* 0,94, dan *recall* 0,75 untuk sentimen positif. Nilai *F1-score* masing-masing adalah 0,88 (positif) dan 0,64 (negatif). Hal ini menunjukkan bahwa model sangat efektif dalam mengenali sentimen positif, namun masih kurang optimal dalam mendeteksi komentar negatif, terutama yang mengandung sarkasme atau ironi. Temuan ini menunjukkan bahwa SVM cocok untuk teks dengan struktur eksplisit namun memerlukan peningkatan saat menangani ekspresi linguistik kompleks.

Daftar Referensi

- [1] N. Dwi Arviana, I. Baidlowi, and M. S. Hidayat, "Pengaruh Strategi Green Marketing, Green Product dan Green Brand terhadap Keputusan Pembelian Produk Kecantikan Ramah Lingkungan," *Journal of Management and Social Sciences*, vol. 3, no. 3, pp. 104–117, Aug. 2024, doi: 10.55606/jimas.v3i3.1437.
- [2] N. A. Ulfah, "Korelasi Rambut Berwarna dengan Stigma Kecantikan Perempuan Urban : Studi Kasus Kota Semarang," *Jurnal Multimedia Dehasen*, vol. 3, no. 4, pp. 307–316, 2024, doi: 10.37676/mude.v3i4.6511.
- [3] O. Wardatus Shalihah, A. Irvie Aranda Alben Susanto, E. Permana, U. Pancasila, and J. Selatan, "Strategi Pengembangan Pemasaran Digital Produk Make-Up Wardah Pada Platform Media Sosial," *Journal of Innovation Research and Knowledge*, vol. 4, no. 11, pp. 8485–8498, 2025.
- [4] I. Komang Dharmendra, I. Made, A. W. Putra, and Y. P. Atmojo, "Evaluasi Efektivitas SMOTE dan Random Under Sampling pada Klasifikasi Emosi Tweet," *Informatics for Educators And Professionals : Journal of Informatics*, vol. 9, no. 2, pp. 192–193, 2024, doi: 10.51211/itbi.v9i2.
- [5] R. C. Larasati, C. Dewi, and H. J. Christanto, "Analisis Sentimen Produk Kecantikan Jenis Moisturizer Di Twitter Menggunakan Algoritma Super Vector Machine," *Jurnal TEKINKOM*, vol. 7, no. 1, pp. 124–134, 2024, doi: 10.37600/tekinkom.v7i1.1243.
- [6] B. A. Maulana, M. J. Fahmi, A. M. Imran, and N. Hidayati, "Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 375–384, Feb. 2024, doi: 10.57152/malcom.v4i2.1206.
- [7] F. M. Herza, B. Rahmat, and M. M. AL Haromainy, "Pengaruh Rfe Terhadap Logistic Regression Dan Support Vector Machine Pada Analisis Sentimen Hotel Shangri-La Surabaya," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 6, pp. 11612–11619, 2024, doi: 10.36040/jati.v8i6.11272.
- [8] B. A. B. Ralahallo, "PERAN MEDIA SOSIAL DALAM MENINGKATKAN PENJUALAN PADA PLAFORM E-COMMERCE DI INDONESIA," *JIRK Journal of Innovation Research and Knowledge*, vol. 5, no. 1, 2025.
- [9] Q. Ayuni, R. Cahya Wihandika, and N. Yulistira, "Klasifikasi Aritmia Dari Hasil Elektrokardiogram Menggunakan Metode Support Vector Machine," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 6, pp. 2163–2170, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [10] G. N. C. Nugroho, "Klasifikasi Gender Berdasarkan Sidik Jari Menggunakan Principal Component Analysis dan Support Vector Machine," *JIEET (Journal Information Engineering and Educational Technology)*, vol. 8, no. 1, pp. 45–53, 2024, doi: 10.26740/jieet.v8n1.p45-53.
- [11] J. Rusman, B. Z. Haryati, and A. Michael, "Optimisasi Hiperparameter Tuning pada Metode Support Vector Machine untuk Klasifikasi Tingkat Kematangan Buah Kopi," *Jurnal Komputer dan Informatika*, vol. 11, no. 2, pp. 195–202, Oct. 2023, doi: 10.35508/jicon.v11i2.12571.

-
- [12] A. R. Damanik, S. Annisa, A. I. Rafeli, A. S. Liana, and D. S. Prasvita, "Klasifikasi Jenis Buah Cherry Menggunakan Support Vector Machine (SVM) Berdasarkan Tekstur dan Warna Citra," in *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA) Jakarta-Indonesia*, 2022. [Online]. Available: <https://www.kaggle.com/moltean/fruits>
- [13] B. Pamungkas, M. E. Purbaya, and D. A. K. Januarita, "Analisis Sentimen Twitter Menggunakan Metode Support Vector Machine (SVM) pada Kasus Benih Lobster 2020," *Journal of Informatics, Information System, Software Engineering and Applications*, vol. 3, no. 2, pp. 10–20, 2021, doi: 10.20895/INISTA.V3i2.
- [14] M. R. B. Keliat and M. Ikhsan, "Komparasi Algoritma Support Vector Machine dan Naïve Bayes pada Klasifikasi Jenis Buah Kurma berdasarkan Citra Hue Saturation Value," *Sistemasi: Jurnal Sistem Informasi*, vol. 14, no. 1, pp. 2540–9719, 2025, doi: 10.18495/generic.v14i1.122.
- [15] T. M. P. Aulia, N. Arifin, and R. Mayasari, "Perbandingan Kernel Support Vector Machine (SVM) Dalam Penerapan Analisis Sentimen Vaksinisasi Covid-19," *SINTECH JOURNAL*, vol. 4, no. 2, pp. 139–145, 2021, doi: 10.31598/sintechjournal.v4i2.762.
- [16] N.A. Irdianto, & R.A. Putri, "Penerapan Algoritma Support Vector Machine Pada Review Aplikasi Tiktok. *Progresif: Jurnal Ilmiah Komputer*, vol. 20, no. 2, pp. 872-885, 2024
- [17] M.R. Pratama, Y.R. Ramadha, & M.A. Komara, "Analisis Sentimen BRLmo dan BCA Mobile Menggunakan Support Vector Machine dan Lexicon Based. *Jutisi: Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, vol. 12, no. 3, pp. 1439-1450, 2023.